DOCUMENT RESUME

ED 161 958 TM 008 037

AUTHOR Sternberg, Robert J.
TITLE Psychometrics, Mathematical Psychology, and Cognition: Confessions of a Closet Psychometrician. Technical Report No. 10.
INSTITUTION Yale Univ., New Haven, Conn. Dept. of Psychology.
SPONS AGENCY Office of Naval Research, Washington, D.C. Psychological Sciences Div.
PUB DATE Jun 78
CONTRACT N00014-78-C-0025
NOTE 50p.; Paper presented at the Annual Meeting of The Psychometric Society (Hamilton, Ontario, Canada, August, 1978)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Cognitive Measurement; *Componential Analysis; Intelligence; *Psychology; *Psychometrics; Reliability; *Research Methodology; *Statistical Analysis; Validity

ABSTRACT

The importance of validity and reliability in the componential analysis of human intelligence is discussed. The roles of construct, internal, external, convergent, discriminant, and ecological validity in componential analysis are outlined. Within-replication (internal-consistency) and between-replication (alternate-forms or test-retest) reliability are also described and illustrated with examples from the author's research on analogical reasoning. The psychometric and multivariate techniques used in componential analysis of intelligence are described by the author, including linear multiple regression, nonlinear multiple regression, canonical regression, principle-axis analysis, confirmatory maximum-likelihood factor analysis, nonmetric multidimensional scaling, and additive clustering. Following a discussion of these procedures, the author describes the evolution of psychometrics and cognitive psychology as separate disciplines. The author contends that cognitive psychologists and psychometricians would benefit from an active interchange, and that psychometricians have technical skills to teach the cognitive psychologist. The first step in an interchange would be to convince the psychologist that these skills are worth learning. Psychometrists can accomplish this by showing cognitive psychologists that psychometrics is responsive to the needs of cognitive researchers. (Author/JAC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED161958

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Psychometrics, Mathematical Psychology, and Cognition

Confessions of a Closet Psychometrician

Robert J. Sternberg

Department of Psychology
Yale University
New Haven, Connecticut 06520

Technical Report No. 10
June, 1978

Approved for public release; distribution unlimited.
Reproduction in whole or in part is permitted for
any purpose of the United States Government.

This research was sponsored by the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract No. N0001478C0025, Contract Authority Identification Number NR 150-412.

TM008 037

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1 REPORT NUMBER: Technical Report No. 10

2 GOVT ACCESSION NO.:

3 RECIPIENT'S CATALOG NUMBER:

4 TITLE (and Subtitle): "Psychometrics, Mathematical Psychology, and Cognition: Confessions of a Closet Psychometrician"

5 TYPE OF REPORT & PERIOD COVERED: Periodic Technical Report (1 Apr 78 - 30 Jun 78)

6 PERFORMING ORG. REPORT NUMBER: Research Report No. 10-78

7 AUTHOR(s): Robert J. Sternberg

8 CONTRACT OR GRANT NUMBER(s): N0001478C0025

9 PERFORMING ORGANIZATION NAME AND ADDRESS:
Department of Psychology
Yale University
New Haven, Connecticut 06520

10 PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
61153N
RR 042-04; RR 042-04-01;
NR 150-412

11 CONTROLLING OFFICE NAME AND ADDRESS:
Personnel and Training Research Programs
Office of Naval Research (Code 458)
Arlington, Virginia 22217

12 REPORT DATE: 30 Jun 78

13 NUMBER OF PAGES: 46

14 MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):

15 SECURITY CLASS. (of this report): UNCLASSIFIED

15a DECLASSIFICATION DOWNGRADING SCHEDULE:

16 DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17 DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18 SUPPLEMENTARY NOTES

Invited address presented to annual meeting of The Psychometric Society, August, 1978, McMaster University, Hamilton, Ontario. This meeting was held in conjunction with the annual meeting of The Society for Mathematical Psychology.

19 KEY WORDS (Continue on reverse side if necessary and identify by block number)

psychometrics, mathematical psychology, cognition, componential analysis, validity, reliability, regression, factor analysis, nonmetric multidimensional scaling, additive cluster analysis

20 ABSTRACT (Continue on reverse side if necessary and identify by block number)

This article presents the confessions of a closet psychometrician. The introduction to the article contains my first public admission of the strange, secret life I have been leading. The remainder of the article is divided into three parts. The first two parts describe how my research on the componential analysis of human intelligence draws upon the disciplines of psychometrics, mathematical psychology, and cognition. In the first part, I describe the relevance of the psychometric constructs of validity and

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

reliability to this research. In the second part, I describe the application of multivariate techniques of regression, factor analysis, nonmetric multidimensional scaling, and additive clustering to the research. The third part of the article explains why I originally became a closet psychometrician, and why I have remained one. I attempt through this explanation to convey what I believe to be the major problem currently facing psychometrics as a discipline.

Abstract

This article presents the confessions of a closet psychometrician. The introduction to the article contains my first public admission of the strange, secret life I have been leading. The remainder of the article is divided into three parts. The first two parts describe how my research on the componential analysis of human intelligence draws upon the disciplines of psychometrics, mathematical psychology, and cognition. In the first part, I describe the relevance of the psychometric constructs of validity and reliability to this research. In the second part, I describe the application of multivariate techniques of regression, factor analysis, nonmetric multidimensional scaling, and additive clustering to the research. The third part of the article explains why I originally became a closet psychometrician, and why I have remained one. I attempt through this explanation to convey what I believe to be the major problem currently facing psychometrics as a discipline.

Psychometrics, Mathematical Psychology, and Cognition: Confessions of a Closet Psychometrician

I had originally intended to use this occasion to come out of the closet--to proclaim publicly, once and for all, that I am and have been for some time a closet psychometrician, at least a part-time one. I had planned, further, to come out of the closet wearing a stylish psychometrician's suit, so that I could be identified readily as a member of the profession. But during the time since I was asked to address you here today, I came to realize that I could not come out of the closet today, or perhaps, ever. The reason for this sad state of affairs is that I no longer have a presentable psychometrician's suit in my wardrobe. The problem, quite simply, is that I've been in the closet for so long, occupying the space that should have been occupied by successively more up-to-date clothing. If I had a presentable suit, I would no doubt be able to solve at least some of the psychometric and mathematical problems I will present to you; but I can't solve any of them, and so will present them to you in the hope that some of you may find them worthwhile to pursue and to solve.

I will divide my presentation into three parts. The first two parts describe how my research on the componential analysis of human intelligence draws upon the disciplines of psychometrics, mathematical psychology, and cognition. In the first part, I describe the relevance of the psychometric constructs of validity and reliability to this research. In the second part, I describe the application of multivariate techniques of regression, factor analysis, nonmetric multidimensional scaling, and additive clustering to the research. The third part of the article explains why I originally be-

came a closet psychometrician, and why I have remained one. I attempt through this explanation to convey what I believe to be the major problem currently facing psychometrics as a discipline.

Psychometric Indices in Componential Analysis[1]

The psychometric indices of validity and reliability play important roles in the componential analysis of human intelligence. These roles will now be described, and illustrated with examples from my research on analogical reasoning (Sternberg, 1977a, 1977b). Comparable analyses have been performed for classification and series completion problems (Sternberg, Note 1; Sternberg & Gardner, Note 2), causal inference problems (Sternberg & Schustack, Note 3), linear syllogisms (Sternberg, Note 4, Note 5), categorical syllogisms (Guyote & Sternberg, Note 6; Sternberg & Turner, Note 7), and conditional syllogisms (Guyote & Sternberg, Note 6).

Validity

This section will deal with several of the many different types of validity, namely, construct validity, internal and external validity, convergent and discriminant validity, and ecological validity.

Construct validity. Construct validity is of signal importance in componential analysis. Indeed, "from a differential viewpoint, componential analysis may be viewed as a detailed algorithm for construct validation, the effort to elaborate the inferred traits (which, in our case, are mental operations) determining test behavior (Campbell, 1960)" (Sternberg, 1977b, p. 65). Other forms of validity to be considered below may be viewed as subordinate to construct validity, in that they are used in the service of the construct validation of a theory or subtheory of intelligence.

Much of my research during the past few years has been devoted to the construct validation of a subtheory of intelligence I call a "unified componential theory of human reasoning" (Sternberg, Note 8). The theory is "unified" in the sense that it attempts to explain within a single theoreti-

cal framework human information processing in a wide variety of reasoning tasks. The theory is "componential" in the sense that the basic unit of information processing in the theory is the component: an elementary information process (Newell & Simon, 1972) that is executed in the solution of some class of problems. This subtheory of intelligence itself comprises hierarchically nested subtheories of reasoning that account for performance on successively more narrow classes of tasks. Theories at each level of the hierarchy include as special cases all subtheories nested beneath them. For example, the theory of analogical reasoning is a special case of a more general theory of induction, which is a special case of the unified theory of reasoning. Components of information processing used in tasks accounted for by subtheories lower in the theoretical hierarchy are claimed also to be used in tasks accounted for by subtheories higher in the theoretical hierarchy. For example, the six components of the theory of analogical reasoning are a subset of the components of the theory of induction, which in turn are a subset of the components of the unified theory of reasoning. Communalities in performance on various kinds of reasoning tasks are explained (in part) by overlap in the components of information processing used to perform these tasks. For example, the components of the theory of analogical reasoning are theorized also to be used in the solution of classification and series completion tasks. Since the unified theory specifies what overlap there should be for various sets of tasks, a major part of the construct validation of the theory is the demonstration that these overlaps do indeed occur, and that others do not. To date, at least, the results of construct validation of the theory have been most encouraging (Sternberg, Note 8).

In componential analysis, construct validation of this and other subtheories of intelligence is guided by a metatheoretical framework according

to which the mental abilities involved in intelligent behavior are viewed as being representable at four successively deeper levels of analysis (Sternberg, Note 9). The first level of analysis is that of the composite task as it appears in standard tests of intelligence. This level may be viewed as the level of the manifest trait or ability. An individual's manifest (first level) ability in analogical reasoning would be measured by his response times and error rates in solving analogies of the form A : B :: C : D. The second level of analysis is that of the subtask: The composite task is decomposed into nested subtasks that require successively less information processing for their completion. This level is an intermediate one that is useful because it often permits isolation of components that, without the use of subtasks, would be experimentally confounded (see Sternberg, 1977b). In the analogies task, a typical subtask would require processing of the C and D terms of the analogy after the subject has been precued with (given as advance information) the A and B analogy terms. The third and fourth levels of analysis may be viewed as levels of the latent trait or ability. The third level of analysis is that of the information-processing component: The subtasks are decomposed into the components of information processing that account for performance on the subtasks. An example of such a component is inference, the process by which the subject discovers and tests the relationship between the A and B terms of the analogy. The fourth level of analysis is that of the information-processing metacomponent: Performance on the components, and particularly on that (constant) component which is required for solution of all experimental manipulations of the composite task under consideration, is controlled by information-processing metacomponents (executive processes). It is via these metacomponents that the subject decides (among other things) what components to use in task solution. In analogical reasoning, for example,

the adult subject presumably decides that inference is needed to solve the analogy. We have collected data, however, that suggest that young children do not always make this decision, relying upon word associations between the C and D analogy terms to solve the problem (Sternberg & Nigro, Note 10).

Internal and external validity. Internal and external validation are used in the construct validation of a componential theory. Internal validation consists of the identification of the (a) basic information-processing components used in composite-task and subtask performance, (b) representation(s) upon which these components act, (c) strategies by which the components are combined, and (d) durations, difficulties, and probabilities of component execution. The validation is internal in that it provides confirmation for a theory of task performance only as the theory relates to performance on the particular task. In the theory of analogical reasoning, for example, (a) six component processes are theorized (b) to act upon an attribute-value representation for information (c) via a strategy in which certain specified components are exhaustive and others self-terminating (d) with durations and difficulties estimated from latency and error data respectively.

External validation consists of the demonstration of the generality of the components, representations, strategies, and parametric values of the components beyond the particular task being studied. The validation is external in that it provides confirmation for a theory of task performance only as it relates to performance on other tasks. In the theory of analogical reasoning, for example, the components, representations, strategies, and parametric values of the components are alleged to be generalizable to series completion and classification tasks as well.

It is common for cognitive research to provide internal but not external

validation of a theory. This validation procedure is inadequate, because it provides no demonstration that any of the properties of performance on the given task are of any interest beyond that particular task. Psychometric research, on the other hand, often provides external but not internal validation of a theory. This validation procedure is also inadequate, because although it may show relations among tasks, it tells the investigator virtually nothing about the internal structure of the task and performance on it.

Convergent and discriminant validity. Convergent and discriminant validation are used in the external validation of a componential theory. Convergent validation consists of the demonstration that identified components of information processing are highly correlated across subjects with external scores with which, theoretically, they should be correlated. Thus, for example, components of analogical reasoning should show high correlations with scores on standardized tests of reasoning ability. Discriminant validation consists of the demonstration that identified components of information processing are uncorrelated across subjects with external scores with which, theoretically, they should not be correlated. For example, components of analogical reasoning should show trivial correlations with scores on standardized tests of perceptual speed. Both convergent and discriminant validation may be assessed either predictively or concurrently. In my own research, I usually give ability tests immediately after the experimental task(s) of interest, so that for all practical purposes, validity is concurrent. Should an attempt ever be made, however, to use component measurements for practical purposes, demonstrations of predictive as well as concurrent validity would be essential.

Ecological validity. Ecological validation consists of the demonstration

that the proposed theory has some kind of practical relevance. Does anything we learn about behavior in the laboratory tell us anything of interest about behavior or potential behavior outside the laboratory? I believe that componential theories can be useful in informing us about and possibly helping to alter behavior in real-world settings. Although we have a training study planned that makes use of the theory of analogical reasoning, the study has yet to be performed. We have done an instructional study employing linear syllogisms, however (Sternberg & Weil, Note 11). In this study, adult subjects were required to make transitive inferences of the kind required by problems such as "John is taller than Pete. Pete is taller than Bill. Who is tallest?" Subjects were divided into three groups. In one (control) group, subjects were not trained to use any particular strategy. In a visualization group, subjects were trained to use a strategy that required visualization of a linear array to solve the problem. In a linguistic (non-visualization) group, subjects were trained to use a strategy that required only linguistic manipulations and no spatial visualization. A major purpose of the study was to determine whether the correlation between performance on transitive inference tasks and spatial visualization could be reduced. If it could, then training low-spatial subjects in the purely linguistic strategy might enable them better to make the transitive inferences required in everyday life. The experiment was successful: The correlation between linear-syllogism response times and spatial ability test scores was statistically significant and highest in the spatially-trained group, significant and intermediate in the untrained group, and nonsignificant and lowest in the linguistically-trained group. Thus, there is some evidence that at least one componential theory, that for linear syllogisms, may be of interest outside of a laboratory setting.

Reliability

This section deals with the application of two types of reliability measurement to componential analysis: within-replication (internal-consistency) reliability and between-replication (alternate-forms or test-retest) reliability. Each type of reliability will be considered in turn.

Within-replication reliability. Within-replication reliability measures the internal consistency of a set of data in which there are no replications. Whereas this type of reliability is usually computed only over subjects in psychometric analyses of task performance, it may be computed both over subjects and over item types in componential analyses of task performance.

When computed over subjects, within-replication reliability serves much the same purpose in componential analyses as in psychometric analyses. Reliability can be computed for scores at any of the four levels of analysis: composite task, subtask, component, or metacomponent. Reliabilities of these scores are of particular importance in assessing strengths of relationships between these internal scores on the one hand, and external scores (such as on standardized ability tests) on the other. Their importance derives from the fact that there can be considerable range in the reliabilities of parameter estimates of duration or difficulty. In my analyses of analogical reasoning, for example, I found that two parameter estimates for durations of component processes generally tended to be much more reliable than the other four. Correlations of the less reliable parameter estimates with external ability tests were consistently lower than those of the more reliable parameter estimates; but correction for attenuation in the parameter estimates suggested that these lower correlations were due primarily to differences in the reliabilities of the parameter estimates. Without this knowledge, I might have attributed the differences in correlation to psychological causes.

When computed over item types, within-replication reliability estimates the upper limit for the proportion of variance in the data that can be accounted for by a model of task performance. If the predictor variables in the model are error-free, then the reliability index (square root of the reliability coefficient) should be used instead of the reliability coefficient. The estimate of the upper limit is only a rough estimate, because most parameter estimation techniques capitalize upon error in the data, so that it is possible for the proportion of variance accounted for to exceed the reliability if error variance is being "accounted for." Nevertheless, the internal-consistency of a set of data provides important information in the interpretation of the goodness of fit of a model. An R^2 of .7, for example, means very different things for two data sets, one having a reliability of .7 and the other having a reliability of .9. In my work on analogical reasoning, the values of R^2 for the preferred model have generally been close to the values of the reliability coefficients, but far enough away for the residual unexplained variance to be statistically significant.

Between-replication reliability. Between-replication reliability measures consistency of sets of data across replications. The distinction between test-retest and alternate-forms versions is of importance only when the same subjects receive both replications, and when these subjects recognize repetitions of item. Because, in componential investigations, subjects often receive large numbers of structurally similar items, they generally do not recognize repetitions of identical items. On the other hand, some types of items, for example, verbal analogies, are almost always remembered. Between-replication reliability, like its within-replication counterpart, may be computed both over subjects and over item types.

When computed over subjects, between-replication reliability indicates

stability of strategy and of relative efficiency in the use of strategy over time, and, possibly, experimental treatments. I have found in my analogies research, for example, that adult subjects tend to use the same strategy for solving analogies over large numbers of trials. But the rank order of subjects with respect to the efficiency with which they use this strategy changes. Although within-replication reliability is high, therefore, between-replication reliability is low. And whereas latency scores for a first session of testing on analogy problems were only poorly correlated with standard tests of reasoning ability, latency scores for a fourth (and last) session of testing were highly correlated with the same reasoning tests (Sternberg, 1977b, Chapter 7). These results demonstrate the necessity of assessing both between- and within-replication reliability, and of modeling separately data collected from subjects at different stages of practice.

Psychometric and Multivariate Techniques in Componential Analysis

In this part of the article, I discuss some uses of psychometric and multivariate techniques in componential analyses of cognition, and particularly, of intelligence. The techniques to be discussed are regression, including linear and nonlinear multiple regression, and canonical regression; factor analysis, including principal axis and confirmatory maximum-likelihood methods; nonmetric multidimensional scaling; and additive clustering.

Regression

Linear multiple regression. Linear multiple regression has played a vital part in both the internal and external validation of the various subtheories of intelligence I have proposed (see Sternberg, Note 8). Each of these two kinds of uses will be considered here.

My primary use of linear multiple regression for internal validation has been in the mathematical modeling of response times and error rates. Typi-

cally, independent variables in these analyses have been numbers of component executions required for solution of various experimental manipulations of the task of interest; the regression weights estimated for these independent variables have then been used to infer the duration or difficulty of each component operation. It is assumed in these analyses that the components contribute additively to total response time and error rate (see Sternberg, 1977a, 1977b).

Overall values of R^2 and root-mean-square deviation (RMSD), and the various F-values associated with them, are useful in comparing the fit of the (a) null model ($R^2=0$) to the proposed model, (b) true model ($R^2=r_{xx}$, the reliability of the data) to the proposed model, and (c) alternative plausible models to the proposed model. F-values for individual parameters (regression weights) are useful in evaluating whether the inclusion of each parameter in the mathematical model can be justified. Nonsignificant parameters may have to be deleted from the model, combined with other parameters, or reconceptualized.

Consider, as an example, research I have done on linear syllogisms (Sternberg, Note 4, Note 5). (Linear syllogisms, it will be recalled, are problems like "John is taller than Bill. Bill is taller than Pete. Who is tallest?") In the research on linear syllogistic reasoning, I pitted three alternative models of transitive inference against each other--a spatial model, according to which the transitive inference linking John to Pete is made by ordering the three terms of the problem in a visualized linear array; a linguistic model, according to which the transitive inference is made by operations performed upon the linguistic deep structures containing the terms, John and Pete;[2] and my own mixed model, according to which the transitive inference is made by a combination of linguistic and spatial operations. The former are used in decoding the premises from the surface structure in which

they are presented to a deep structure in which they can be operated upon; the latter are used in recoding the deep structure into a spatial array that is used to relate John and Pete. Linear multiple-regression analyses of latency data from a series of experiments revealed that (a) all of the models were superior to the null model, (b) none of the models could have been the true model, since none accounted for all of the reliable variance in the data, and (c) the mixed model was consistently superior to the alternative spatial and linguistic models. The parameters of the mixed model were generally statistically significant, and were replicable across a variety of subjects and experimental conditions.

My primary use of linear multiple regression for external validation has been in the mathematical modeling of reference ability scores. These scores, to be discussed further later, are factor scores for constructs such as inductive reasoning, spatial visualization, and the like. The purpose of the modeling is to show that individual differences in these constructs can be explained in terms of individual differences in the information-processing components theorized to underlie the constructs.

Consider again the research on analogical reasoning (Sternberg, 1977b). External validation was used to demonstrate that the proposed components of analogical reasoning accounted for much of the variance in an inductive reasoning reference ability score (convergent validation), but little of the variance in a perceptual-speed reference ability score (discriminant validation). Subjects' factor scores were predicted from their latency component scores on the various components of information processing considered simultaneously. The obtained data were consistent with the proposed relationships.

The most serious problem associated with the use of linear multiple regression in research on cognition may well be that of multicollinearity among independent variables. As the correlations among pairs of independent variables increase, the interpretability and replicability of the regression coefficients decreases. Suppressor effects become increasingly common, so that one often obtains negative regression coefficients in situations where only positive ones make theoretical sense. Overall, it becomes increasingly difficult to assess the independent contribution of each independent variable to prediction of the dependent variable (Darlington, 1968).

There are two routes I am aware of for solving the problem of multicollinearity. The first is to design experiments so that independent variables that are naturally correlated (in the population) are rendered orthogonal (in the sample). Statistical problems associated with pseudo-orthogonalization have been forcefully demonstrated by Humphreys and Fleishman (1974), and the most obvious conceptual problem is that one runs the risk of obtaining results that do not reflect the natural situation of interest, but rather an artificially contrived situation of no interest. The second route is that of statistical procedures that attempt to give robust regression weights despite the correlations among independent variables, for example, the jackknife (Mosteller & Tukey, 1968, 1977) and ridge regression (Hoerl & Kennard, 1970a, 1970b, 1976; Price, 1977). The second route seems to be preferable to the first route, in that it attempts to deal with the problem of multicollinearity rather than pretending that it's not there. The major problem with the second route is that statisticians and psychometricians know relatively little about the properties of these procedures in practical applications, and many cognitive psychologists are unaware

that they even exist. What cognitive psychologists need, therefore, is first, more research on the uses and usefulness of these techniques in experimental settings, and second, increased education about and accessibility of the techniques. The research of Wainer (Wainer, 1976a, 1976b, 1978; Wainer & Thissen, 1975) is a step toward fulfilling at least the first need. And I hope that within the next few years, computer programs that implement these techniques will become more readily accessible. I suspect, though, that many cognitive psychologists first will have to be educated in the use of regression as a tool in research: Many of them still harbor the illusion that to be a true experiment (in some Platonic sense of the word), a study must use a factorial design that is immediately susceptible to analysis of variance.

A second problem associated with the use of linear multiple regression is the large numbers of observations needed from each subject to obtain stable estimates of regression parameters for individual subjects. If one takes parameter estimates for individual subjects seriously, as I must when I correlate these parameters with each other and with scores on reference ability tests, then one must assure oneself that the data from which the parameters are estimated are reasonably reliable. This assurance requires many replications of individual data points for each subject. Utter exhaustion in testing subjects often then leads to relatively smaller numbers of subjects in a given experiment. But these smaller numbers of subjects, in turn, mean that correlational analyses interrelating parameters of mathematical models to each other and to external criteria will be lacking in power. As a result, interesting relationships may go undetected because they are too weak to be spotted with the low-power test used. What can be done so that for a given number of subject hours available, an optimum tradeoff can be achieved between

numbers of subjects (and hence power of correlational and other tests) and numbers of observations per subject (and hence reliability of parameter estimates for each subject)?

Some data points tend to be more reliable than others. If, for example, the item type corresponding to one data point requires a relatively large number of information-processing components in its solution, and the item type corresponding to another data point requires a relatively small number of components in its solution, the second item type will probably require fewer replications than the first in order to achieve the same accuracy of measurement. Techniques would be welcome, therefore, that would permit estimation before testing of the numbers of observations needed for each data point to be measured at a given level of accuracy. Prior information to make this determination would include (a) a theoretical account of what operations are used in the solution of various item types, and (b) estimates of the parameters corresponding to the durations or difficulties of these operations, plus the standard errors associated with these parameter estimates. Alternatively, it might be possible to adopt some of the techniques of adaptive testing (Lord, Note 12; Weiss, Note 13; Weiss & Betz, Note 14, Note 15) to information-processing measurements in much the same way they have been adopted to psychometric measurements. Subjects would then receive test items that optimize measurement of their individual parameters. The development I have seen that is closest to this goal is in computer-assisted instruction for foreign-language learning (Atkinson, 1972).

A third problem I have encountered is a strictly practical one--the unavailability of constraints for parameters in standard linear multiple regression programs such as SPSS (the Statistical Package for the Social Sciences).

Some computer programs, such as BMD (the Biomedical package) will force a zero intercept during the regression, but will allow no other, possibly more sophisticated, constraints. Often in information-processing research, however, one would like to constrain parameters to be nonnegative, to sum to a certain number, or to bear other single or multiple relations to each other. If constraints were built into linear regression programs, as they are into at least some nonlinear regression programs, much greater flexibility could be achieved in the modeling of cognitive processes.

Nonlinear multiple regression. My colleagues and I have used nonlinear regression in the testing of response-choice models for both induction tasks (Sternberg & Gardner, Note 2) and deduction tasks (Guyote & Sternberg, Note 6; Sternberg & Turner, Note 7). Nonlinear regression has been used only for internal (and not external) validation of the models.

Consider, as an example, a problem commonly used to measure deductive reasoning ability, the categorical syllogism. A typical categorical syllogism is "All B are C. All A are B. Which of the following conclusions is logically valid? All A are C; Some A are C; No A are C; Some A are not C; None of these conclusions." Subjects are instructed to choose the best of the possible conclusions, should more than one be logically valid (as in the above case, where both "All A are C" and "Some A are C" are acceptable).

According to our proposed transitive-chain theory of syllogistic reasoning (Guyote & Sternberg, Note 6), syllogistic reasoning can be decomposed into four global stages of information processing: encoding, during which premise information is read and translated into a canonical symbolic representation expressing all possible set relations corresponding to each premise; combination, during which the symbolic representations are integrated via what we call "transitive chains;" comparison, during which the combined representation is

compared to the possible conclusions; and response, during which the subject communicates the chosen conclusion (or the belief that no conclusion is valid).

According to the transitive-chain theory, encoding and response are executed without error. There are thus no parameters of response choice associated with these stages of information processing. Errors during the combination stage of syllogistic reasoning are theorized to result from limitations in the ability of working memory to hold all possible combinations of encoded set relations. Four parameters--p_1, p_2, p_3, p_4--are used to represent probabilities of combining exactly one, two, three, or four pairs of set relations respectively. It is assumed that subjects never combine more than four pairs of set relations (of sixteen possible in the most complexly represented syllogism). Errors during the comparison stage of syllogistic reasoning are theorized to result from simplifying heuristics subjects use to facilitate selection of a conclusion for their final combined representation. Three parameters--β_1,β_2, c--are used to represent probabilities that three possible heuristics (described in Guyote & Sternberg, Note 6) are used.

Four alternative models of syllogistic reasoning were compared to the transitive-chain model for their ability to account for response-choice data with different sets of subjects and over a fairly wide variety of experimental conditions. It was found that (a) all of the models were superior to the null model ($\eta^2=0$),(b) none of the models were comparable to the true model ($\eta^2=r_{xx}$), and (c) the transitive-chain model was clearly superior to the alternative models in its ability to account for the observed response-choice probabilities in each (of six) data sets.

Parameter estimation was done using the BMDP3R program from the Biomedical P

series. My requests for developments in nonlinear multiple regression are both of an applied nature.

First, the output provided by these programs seems generally to be inadequate. They usually provide neither a value of η^2 nor a value of root-mean-square deviation (RMSD). Although these values usually can be calculated from the output that is given, the calculations could be performed much more rapidly by computer than by hand or by calculator. Moreover, the programs generally give no indication of error of estimate associated with the various parameters. Users need to know, however, how much faith they can place in the various parameter estimates they receive. Finally, output from the programs is often scantily labeled and difficult to read, rendering more difficult the already difficult job of interpreting the various numbers that the computer has printed out.

Second, further refinements are needed in defining rational starting values and sensible step sizes. I have found in these programs that the starting values one uses can have a substantial impact upon the final parameter estimates one obtains, and that the step sizes do not always appear to be optimal for convergence upon absolute minima. In fact, the programs I have used seem to be quite susceptible to local minima. Difficulties in defining rational starting configurations and variable step sizes have also confronted multidimensional scalers over the years, but psychometricians such as Shepard (1962a, 1962b), Kruskal (1964a, 1964b), and Young (Young, 1970; Young & Torgerson, 1967) have developed ways of solving these problems in a reasonably satisfactory manner. The KYST-2 computer program combines many of the best features of this earlier work. I would like to see the same developments in nonlinear regression, where developments for handling these problems seem to have lagged.

Canonical regression. Canonical regression has of yet received little use in cognitive psychology, although I believe that its use will increase as its usefulness is recognized. The major use to which I have put canonical regression is in the simultaneous modeling of response-time and error data (Sternberg, 1977b; Sternberg, Note 5). As things have stood, ways of handling error data when modeling response times, and of handling response times when modeling error rates, have been unsatisfactory. The overwhelmingly common tendency has been to attempt to model one and either to ignore the other or to attempt to render it irrelevant. Thus, in experiments where response time is the major dependent variable, an effort is almost always made to minimize error rates; in experiments where error rate is the dependent variable, response time usually is just ignored.

Canonical regression seems to provide one way of taking into account both response times and error rates simultaneously. The two are used jointly as dependent variables, and are predicted from a set of independent variables that is theorized to give rise to both latency and error of execution. The assumption must be made that the sources of both latency and error are potentially the same, and additive (see Sternberg, 1977b, for details). The important question then becomes one of whether modeling the two together provides more information than could be obtained simply by looking at one or the other alone.

Consider two brief examples from my research on analogies and linear syllogisms. Modeling of both schematic-picture (People Piece) and geometric analogy data via canonical regression revealed two canonical variates (Sternberg, 1977b, Chapters 7 and 9). The first was essentially identical to solution time. Canonical variate scores for error rates were correlated fairly highly with this variate, but not nearly so highly as were the canonical variate

scores for solution times. Error rate thus appears to be an imperfect, or imprecise measure of whatever it is that solution time measures. The usual moderately high correlations across item types between solution times and error rates can be attributed to this relationship between the two variables. The second canonical variate, which was statistically significant for both types of analogies, represented that part of error rate that is independent of solution time. The loadings of the independent variables on this variate were quite different from those on the first variate, as would be expected. But the pattern of loadings was quite consistent across analogy contents: Self-terminating operations contributed heavily to the prediction of the dependent variate, whereas exhaustive operations did not. Thus, it appears that at least in analogical reasoning, self-terminating operations may be largely responsible for that aspect of error rate that is independent of solution time. Converging evidence for this conclusion has been found in work on the development of analogical reasoning (Sternberg & Rifkin, in press): Children become progressively more nearly exhaustive in their analogical-reasoning operations as they grow older, and at the same time, their error rates on analogies decrease dramatically.

I have previously reported what at first appeared to be a curious conflict in the literature on linear syllogisms (Sternberg, Note 5). Certain data sets, including my own, seemed to support my proposed mixed model of transitive inference, whereas a smaller but nontrivial number of data sets seemed to support an alternative linguistic model (Clark, 1969a, 1969b). Canonical regression and other analyses helped reveal the source of these conflicts in the literature. A fundamental difference between data sets supporting the two models was in whether the dependent variable was solution time or error rate: Solution-time data tended to support the mixed model,

error data, the linguistic model. Canonical regression revealed that certain parameters of the mixed model predicted the first (solution time) but not the second variate (error rate independent of solution time), whereas a single parameter of the linguistic model predicted the second but not the first variate. Both models were incomplete, therefore. The results of the canonical regression suggested what would be needed in order to construct a full model of transitive inference that successfully predicted both solution time and error rate.

Users of canonical regression doing substantive research have at least three immediate needs. The first is for a better understanding of the kinds of substantive questions for which canonical regression is and is not a useful method of data analysis. Cognitive psychologists confronted with multivariate data often do whatever they can to analyze their data in a univariate fashion. I believe they take this route because of their ignorance regarding how multivariate techniques could help them answer questions that univariate techniques cannot answer. Education in the uses of canonical regression is therefore a must. The second need is for more information regarding alternatives to unrotated solutions. I have found that unrotated solutions in a canonical regression tend to result in canonical coefficients showing trends very similar to those shown by the pattern coefficients in an unrotated factor solution: a general factor followed by a series of bipolar factors. This pattern of canonical coefficients may not be the most interpretable one. But rotation in canonical regression is a more serious matter than in factor analysis, because it destroys the orthogonality of the variates. We therefore need to know more about what the options in rotation are, and what kinds of effects they can be expected to have on our data. Finally, canonical regression programs should be written to provide much more information than they

presently do. Generally, they do not permit the option of rotation, nor do they compute canonical variate scores. I have found, however, that the correlations of the canonical variate scores with the dependent and independent variates are almost always more intepretable than are the canonical coefficients. The programs should also provide standard errors of individual coefficients, and the option of entering the independent variables (as defined by the user) in a stepwise fashion.

Factor Analysis

Principal-axis analysis. I have used principal-axis analysis for two very different purposes in my research, and I would like briefly to describe both of these uses here.

The first use has been in my research on intelligence. In the past, factors have often been viewed as source traits or latent traits, in other words, as the underlying dimensions along which individuals differ (Cattell, 1971; Guilford, 1967). I have previously stated why I believe this neither is nor could be the case (Sternberg, 1977b, Chapter 2). In componential analysis, factors are viewed instead as constellations of mental abilities that are organized by patterns of variation across individuals.[3] Factors provide a useful way of reorganizing data at any of the four levels of mental abilities, in order to understand the organization of individual differences at that level.[4] Factors do not provide an additional level, but rather, a differing perspective on a given level. I refer to the constellations of components or metacomponents entering into the factors at a given level as reference abilities.

Consider how we might interpret a factor analysis of a battery of tests of what Horn and Cattell refer to as fluid intelligence (Horn, 1968; Horn & Cattell, 1966). The factors we identify will depend largely upon the rota-

tion we choose to perform (Sternberg, 1977b). The choice is a matter of convenience. One possible pattern, which would be likely to emerge from an unrotated solution, is a general factor, followed by group factors, followed by specific factors. I would interpret the general factor as constituted of those components and metacomponents that are relevant in performance on all of the fluid ability tests; the group factors would comprise those components and metacomponents limited to groups of tests; and the specific factors would comprise those components and metacomponents that are specific to single tests. As mentioned earlier in the article, attempts to account for factor scores by component scores via multiple regression have been quite successful (Sternberg, 1977b).

The second use of principal-axis analysis has been in research on metaphor (Sternberg, Tourangeau, & Nigro, in press; Tourangeau & Sternberg, Note 16, Note 17). This use of factor analysis is based upon the concept of the semantic differential (Osgood, Suci, & Tannenbaum, 1957). Subjects were asked to rate each of 20 terms within each of 8 semantic domains on each of 21 scales, such as warlike-peaceful, noble-ignoble, and strong-weak. We hoped in this way to obtain for each of the eight domains (U.S. historical figures, modern world leaders, mammals, birds, fish, airplanes, land vehicles, and ships) a set of two dimensions (prestige and aggression) that were at least roughly correspondent in each case. We were successful in this regard: Correlations between the loadings of the adjective pairs on dimensions we believed either to correspond or not to correspond were high and low respectively. Since different subjects supplied ratings for each of the domains, there was thus some evidence of between-subject as well as between-domains consistency in the dimensions along which the various domains are perceived. Other subjects rated the eight domain names on each of the adjective scales, and these results

were also factor analyzed. Three factors were obtained, corresponding roughly to three types of domain content (types of people, types of animals, types of vehicles).

The basic idea motivating these analyses is that each of the factor-analyzed domains can be viewed as a local subspace of the hyperspace obtained by factor analyzing the domain names themselves. Thus, each point in the higher-order hyperspace maps into a whole lower-order local subspace. For example, modern world leaders is a point in the hyperspace, but it is also a local subspace in its own right. Since the dimensions of the local subspaces are correspondent, and since factor analysis standardizes the distances within each domain, one can imagine direct comparison of point locations within various local subspaces. For example, wildcat and ICBM may be said to be correspondent if their coordinate values within their respective local subspaces are the same. If the coordinate values are not the same, the degree of noncorrespondence can be measured by what we call the Euclidean superimposed within-subspace distance. We use the term "superimposed" to call attention to the fact that when the distance between the two points is computed, it is computed as though the two domains from which the terms come were superimposed. Now, one can also compute Euclidean distances within the hyperspace. Distances within the hyperspace are actually distances between domains, which are represented by the local subspaces. We therefore refer to within-domain distance as between-subspace distance.

Consider a metaphor such as, "A wildcat is an ICBM among mammals." Our basic theory of metaphor is that a metaphor is comprehensible to the extent that both the superimposed within-subspace distance and the between-subspace distance between tenor (wildcat) and vehicle (ICBM) are small. In other words, if wildcat and ICBM are at nearly correspondent points in their respec-

tive local subspaces, and if the local subspaces are near each other (i.e., their separation within the hyperspace is small), then the metaphor will be readily comprehended. A metaphor is aesthetically pleasing to the extent that the superimposed within-subspace distance is small, but the between-subspace distance is large. In other words, if the two terms are at nearly correspondent points in their respective local subspaces, and if the local subspaces are far apart (i.e., their separation within the hyperspace is large), then the metaphor will be judged as good (or apt). Note that smaller superimposed within-subspace distance works in favor of both comprehensibility and aesthetic pleasingness, but that smaller between-subspace distance works in favor of comprehensibility but against aesthetic pleasingness. Preliminary data provide some support for the theory, although it is too early to tell whether all aspects of it will be confirmed.

Factor analysis has been extensively investigated by psychometricians, and I suspect that the research that still needs to be done on it is less pressing than the research that needs to be done on other techniques. Nevertheless, I see four directions of research that might be helpful to cognitive psychologists. Any such research must of course be communicated to cognitive psychologists in a way that makes clear its usefulness to them.

First, I believe multimode, or interbattery factor analysis (Tucker, 1958) could be exploited by cognitive psychologists if they understood it better (and if more were known about its psychometric properties). Researchers interested in integrating the psychometric and cognitive approaches to intelligence sometimes wish to relate psychometric types of measures to cognitive ones. Hunt, Lunneborg, and Lewis (1975), for example, have sought to do this through conventional factor analysis, with somewhat disappointing results. Psychometric measures tended to group together into their own factor, and cog-

nitive measures into their own factors. Interbattery factor analysis would have prevented this not unpredictable outcome, and might better have pointed out the salient relationships between the psychometric and cognitive measures.

Second, I believe we need to know more about possible uses of factor analysis in multiple and canonical regression analysis. With large numbers of independent variables, or with highly correlated ones, regression weights can become difficult to interpret, and factor analysis provides a way of reducing rank and making outcomes of regression analyses more interpretable. The research of Skinner (1977, 1978) and Tucker (1973) is the kind we need in order to obtain information about the potential uses of factor analysis in regression.

Third, possible uses of factor analysis in exploring intra- as well as inter-item structure need to be examined. Cognitive psychologists are generally more interested in knowing about intra-item processes than in knowing about inter-item structure. Could factor analysis help shed light on these intra-item processes, perhaps by being applied to a set of items, all of which require the same processes,but none of which require the same numbers of these processes? This is the kind of data set I analyzed by multiple regression in my analogies research (Sternberg, 1977a, 1977b). My initial attempts at factor analyzing these data were aborted by my move from Stanford to Yale, and never resumed. If there are individual differences in all of the processes identified by the multiple regression, however, it would seem that a factor analysis of the intercorrelations between all pairs of item types should reveal the same processes. In other words, the component processes that generate differences in response latencies for the various item types should also generate differences in response latencies for the various subjects solving the item types.

Fourth and finally, we need to know more about the potential of Procrustean rotation (Schönemann, 1966) as a tool in cognitive research. The findings of Horn (Horn, 1967; Horn & Knapp, 1973) and of Humphreys (Humphreys, Ilgen, McGrath, & Montanelli, 1969) have given Procrustean rotation something of a bad name. Their research suggests, however, not that Procrustean rotation is intrinsically bad, but that researchers need to be cautious in its use, and to appreciate fully its properties, in particular, its susceptibility to capitalization upon chance. If its limitations do not turn out to be too debilitating, then Procrustean rotation properly used might provide a useful means of testing alternative cognitive theories. Our present state of knowledge regarding Procrustean rotation does not leave a great deal of room for optimism in this regard, but neither does it suggest that Procrustean rotation is useless. We may simply not yet know enough about its properties to use it effectively.

Confirmatory maximum-likelihood factor analysis. I have not yet used confirmatory maximum-likelihood factor analysis (Jöreskog, 1969; Jöreskog & Lawley, 1968) in my own research, but Frederiksen's (Note 18) brilliant use of one variant of this technique, analysis of covariance structures (Jöreskog, 1970), persuades me that confirmatory maximum-likelihood analysis could play a major role in future investigations of various aspects of intelligence. Frederiksen was interested in the components of reading performance, and by a clever combination of psychometric and cognitive techniques was able to isolate factors that seem at least to represent various stages in silent reading. Anything that can be done to further our knowledge about confirmatory maximum-likelihood techniques, and to communicate this knowledge to cognitive psychologists, would be a most useful contribution indeed.

Nonmetric Multidimensional Scaling

Since its introduction by Shepard (1962a, 1962b) and its further development by Kruskal (1964a, 1964b), Young (Young, 1970; Young & Torgerson, 1967), and Guttman (1968), nonmetric multidimensional scaling has played a more important role in cognitive psychology than have most multivariate techniques. Perhaps this is because the originator of the method, Shepard, is himself a cognitive psychologist, and has provided a number of excellent examples of how the technique can be applied to cognitive research.

We have used nonmetric multidimensional scaling in an extension of Rumelhart and Abrahamson's (1973) theory of analogical reasoning to other forms of inductive reasoning (Sternberg & Gardner, Note 2). Rumelhart and Abrahamson used Henley's (1969) animal-name space as a basis for testing a proposed theory of reasoning by analogy. According to the theory, the terms of an analogy composed of animal names can be represented in a space of animal names as a parallelogram. The A and B terms of the analogy are related by a mental vector extending from A to B; the C and ideal (I) terms are similarly related. The vectors relating A to B and C to I are theorized to be parallel. Unfortunately for subjects solving analogies, there will almost never be an animal name at exactly the location corresponding to I. So, given a choice among possible completions, subjects must use some kind of decision rule to choose which of several answer options is best. Rumelhart and Abrahamson proposed the applicability of Luce's (1959) choice axiom to this situation, and were able to make quantitative predictions about choice probabilities by further assuming that the probability of choosing an alternative, X_1, as best, is an exponentially decreasing function of the distance of that alternative from I. Making relatively few assumptions about the nature of the data and the choice process operating upon it, Rumelhart and Abrahamson were able to

obtain excellent fits of their model to observed response-choice probabilities.

In one of our two experiments, we presented subjects with animal-name analogies, series completions, and classifications. In each task, subjects had to rank order four answer options for goodness of fit. In the analogies, subjects had to figure out how well each of the answer options completed a problem such as RAT : PIG :: GOAT : _____ (A) CHIMPANZEE, (B) COW, (C) RABBIT, (D) SHEEP.[5] In the series completion problems, subjects were presented with the first two terms of a series, and had to continue the series: RABBIT : DEER : _____ (A) ANTELOPE, (B) BEAVER, (C) TIGER, (D) ZEBRA. In the classification problems, subjects were presented with three animal names, followed by four options. Subjects had to decide how well each of the four options fit with the first three terms: MOUSE, CHIMPANZEE, CHIPMUNK, (A) GORILLA, (B) RAT, (C) SQUIRREL, (D) ZEBRA.

It was theorized that in each of the three tasks, subjects would employ a somewhat different strategy. These different strategies, however, would be aimed at a common goal, the discovery of an ideal point at which an optimum answer would be located. Subjects would then use the decision rule proposed by Rumelhart and Abrahamson to rank order the four answer options for goodness of fit to the ideal point. A single exponential parameter was estimated from the response-choice data for each task. The values obtained from the three tasks were remarkably similar. Moreover, the identical mathematical model provided an excellent fit to the observed response-choice probabilities in each task. It thus appears that the Rumelhart-Abrahamson theory of response choice in analogical reasoning can be extended to response choices in at least two other inductive reasoning tasks as well.

Although nonmetric multidimensional scaling has seen more use in cognitive research than has practically any other multivariate technique, I am

less sanguine about its future in cognitive research than I am about any of the other techniques I have discussed. The reason for my pessimism is the present limitation in the applicability of nonmetric multidimensional scaling. Current uses of the technique require highly constrained and artificial stimulus spaces. For example, the theory of response choice in analogical reasoning, and its extension to other forms of induction, can be easily applied to a well-defined stimulus domain such as animal names; but the theory seems much less readily applicable to the ill-defined domains that are common in everyday experience. Unless multidimensional scaling can be shown to be useful in these domains as well, I doubt that it will maintain its prominent role in cognitive psychology. Although multidimensional scaling can be applied to any matrix of correlations that factor analysis can be applied to, its advantages over factor analysis remain to be demonstrated, especially with recent developments in nonmetric factor analysis (Kruskal & Shepard, 1974).

One line of research that at one time looked promising was the comparison of different Minkowski r-metrics for the processing of distance information in various tasks. Shepard (1964), Arnold (1971), and others presented evidence that under certain circumstances, subjects might prefer either the city-block metric of $r=1$ (Shepard) or the dominance metric of $r=\infty$ (Arnold) to the standard Euclidean metric of $r=2$. Shepard (1974) has pointed out difficulties in the comparison of r-metrics on the basis of relative levels of stress, however, and it is no longer clear just how different r-metrics can be validly compared. This potentially interesting line of research will be without a future unless some clearly valid way of making these comparisons is found.

Much less is known about the differential properties of various kinds of rotations in multidimensional scaling than is known about these properties in factor analysis. The multidimensional scaling programs I have used have had inadequate provisions for rotation, leaving scaling solutions either with axes in an arbitrary position or in a principal-components position. Because of the importance of rotation to the interpretation of scaling as well as factor-analytic solutions, more needs to be known about the properties and utilities of various kinds of rotations in nonmetric multidimensional scaling.

Additive Clustering

Additive (overlapping) clustering has been an option for researchers for a number of years (Jardine & Sibson, 1971), but it is only recently that the development of algorithms for additive clustering has reached a point where it seems that additive clustering programs will soon be readily available for use by cognitive psychologists (Carroll, 1976; Shepard & Arabie, Note 19; Arabie & Carroll, Note 20). I have used an additive clustering of the Henley (1969) animal-names space performed by Arabie and Rips (and reported in Shepard & Arabie, Note 19) as a means for providing further tests of my information-processing theory of analogical reasoning (see Sternberg, 1977b, Chapter 10).

An attempt was made to compare how well two different sets of independent variables could predict the differential difficulties of various animal-name analogies. One set of independent variables was based upon a spatial representation for information; the other was based upon an additive clustering representation for information. The independent variables formed from the spatial representation were coordinate values for particular analogy terms (used to measure difficulty of an attribute-identification component) and distances between coordinate values for pairs of analogy terms (used to measure

difficulty of certain attribute-comparison components). Each of the three dimensions in the Henley animal-names space was considered separately, since it was found that the dimensions varied in their abilities to predict item difficulty. In the additive clustering representation, easiness of attribute-identification was measured by the number of overlapping clusters in which a given term appeared, and easiness of attribute-comparison was measured by number of overlapping clusters in which two terms appeared together. The ideas motivating these independent variables were, first, that the greater the number of clusters in which a term appeared, the more likely it was that an attribute would be encoded that would later be relevant in comparison, and second, that the greater the number of clusters in which two terms appeared together, the more likely it was that at least some communality would be found between the two terms. The results of the original study, and a replication of it (Sternberg & Gardner, Note 2), supported the additive clustering representation over the spatial one as a means of predicting item difficulty.

The major problem that has confronted potential users of additive clustering is the lack of accessible software for doing it. This situation will be remedied by the exportation of Arabie and Carroll's (Note 20) MAPCLUS program. This program could serve to revitalize the clustering literature in cognitive psychology, where I believe that hierarchical clustering has sometimes been extended to situations in which it is less appropriate than an additive model, simply because of the much greater accessibility of hierarchical clustering programs. Users will soon have both options readily available, and will be able to choose between them on theoretical rather than strictly pragmatic grounds.

How I Became a Closet Psychometrician, and Why I Remain One

I promised, finally, to confess why I first became a closet psychometrician, and why I have remained one. Obviously, I could blame myself for leading this bizarre private life, although I would much prefer to pin the blame on some external target. Alternatively, I could blame certain psychometricians, mathematical psychologists, or cognitive psychologists, or the three groups in general; but culpability lies elsewhere. It lies, I believe, in the ways in which psychometrics and cognition happen to have evolved as psychological disciplines. Each has pretty much gone its own way. In the first half of the twentieth century, the disciplines of cognition (in particular, intellectual cognition) and psychometrics thrived in a symbiotic relationship: Each informed the other. Research on cognition suggested important psychometric problems to be solved, and the solutions to these problems were fed back into the study of cognition. Many of the great early psychometricians--Spearman, Thomson, Thurstone, to name just a few--also maintained active substantive research programs. Even today, I suspect, past presidents of The Psychometric Society contain among them a disproportionate number of psychometricians with strong substantive interests. But the symbiotic relationship between the studies of cognition and psychometrics has fallen by the wayside. Psychometrika today has become, by choice, a purely methodological journal, and much of the research reported in it bears only the most peripheral relationship to substantive concerns. Many if not most psychometricians seem to have little knowledge of or interest in cognition, and many cognitive psychologists are only dimly aware of what psychometrics is. Both sides lose: Cognitive psychologists lose the opportunity to exploit tools that I believe and have attempted to demonstrate can be most useful to them in their research; psychometricians lose touch with what might be really important problems, and risk retreating in their research to esoterica.

I realize that there is another side to the coin--the danger that too close a relationship between two disciplines will lead to the assimilation of one into the other. Consider, for example, the contrast between developments in mathematical psychology and those in psychometrics. The interests of mathematical psychologists have changed with the substantive winds: Emergent mathematical methodologies have very much reflected prevaling substantive concerns. Thus, stochastic modeling was refined in the 1960's to meet the needs of cognitive and other psychologists interested in probabilistic models of learning and concept formation; and regression modeling has been refined in the 1970's to meet the needs of cognitive psychologists interested in information-processing models of reading, reasoning, and the like. On the one hand, I see this path of development as a healthy one. On the other hand, I see something of an identity crisis: The separate identity of mathematical psychology as a discipline has become murky indeed. I recall, for example, Bill Estes facetiously commenting in a recent invited address to what was once called the Mathematical Psychology Group that his address would actually have some mathematics in it.

I would like to think that there is a middle road, and that psychometrics as a discipline is capable of moving toward it. Psychometrics has a great deal to offer cognitive psychology, and I have tried to show in this article how psychometrics has informed and, I hope, enriched my own research. Cognitive psychology also has a great deal to offer psychometrics, and I have tried to show in this article some of the problems that I as a cognitive psychologist (and a closet psychometrician, but don't tell anybody) would like to see solved. If anyone is to start finding the middle road where the two disciplines can interact with each other, I suspect it will be the members of The Psychometric Society rather than those of The Psychonomic Society. Or per-

haps it will be those few psychologists who are members of both and mean it. Psychometricians have the technical skills to teach to the cognitive psychologists, but they will first have to convince them that these skills are worth learning. They will be able to do so when they show them how psychometric techniques can be applied to cognitive research, and how current research in psychometrics is in at least some ways responsive to the needs of cognitive psychologists. When and if this day comes, I will come out of the closet, and indeed, I'll have no choice, because there will no longer be any closet to hide in. And that is as it should be.

Reference Notes

1. Sternberg, R. J. Components of inductive reasoning. Manuscript in preparation, 1978.

2. Sternberg, R. J., & Gardner, M. K. A unified theory of inductive reasoning in semantic space. Manuscript submitted for publication, 1978.

3. Sternberg, R. J., & Schustack, M. W. Components of causal inference. Manuscript in preparation, 1978.

4. Sternberg, R. J. Representation and process in transitive inference. Manuscript submitted for publication, 1978. Also NR 150-412 ONR Technical Report No. 12. New Haven: Department of Psychology, Yale University, September, 1978.

5. Sternberg, R. J. A proposed resolution of curious conflicts in the literature on linear syllogisms. Paper presented at VIIIth International Symposium on Attention and Performance, Princeton, August, 1978. Also NR 150-412 ONR Technical Report No. 8. New Haven: Department of Psychology, Yale University, June, 1978.

6. Guyote, M. J., & Sternberg, R. J. A transitive-chain theory of syllogistic reasoning. Manuscript submitted for publication, 1978. Also NR 150-412 ONR Technical Report No. 5. New Haven: Department of Psychology, Yale University, April, 1978.

7. Sternberg, R. J., & Turner, M. E. Components of syllogistic reasoning. Manuscript submitted for publication, 1978. Also NR 150-412 ONR Technical Report No. 6. New Haven: Department of Psychology, Yale University, April, 1978.

8. Sternberg, R. J. Toward a unified componential theory of human reasoning. Manuscript submitted for publication, 1978. Also NR 150-412 ONR Technical Report No. 4. New Haven: Department of Psychology, Yale University,

April, 1978.

9. Sternberg, R. J. The nature of mental abilities. Manuscript submitted for publication, 1978. Also NR 150-412 ONR Technical Report No. 9. New Haven: Department of Psychology, Yale University, June, 1978.

10. Sternberg, R. J., & Nigro, G. The development of verbal relations in analogical reasoning. Manuscript submitted for publication, 1978.

11. Sternberg, R. J., & Weil, E. An aptitude-strategy interaction in the solution of linear syllogisms. Manuscript submitted for publication, 1978. Also NR 150-412 ONR Technical Report No. 13. New Haven: Department of Psychology, Yale University, September, 1978.

12. Lord, F. M. Tailored testing, an application of stochastic approximation. NR 150-303 ONR Technical Report. Princeton: Educational Testing Service, February, 1971.

13. Weiss, D. J. Strategies of adaptive ability measurement. Research Report No. 74-5. Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota, 1974.

14. Weiss, D. J., & Betz, N. E. Ability measurement: Conventional or adaptive? Research Report No. 73-1. Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota, 1973.

15. Weiss, D. J., & Betz, N. E. Empirical and simulation studies of flexilevel ability testing. Research Report No. 75-3. Minneapolis: Psychometric Methods Program, Department of Psychology, University of Minnesota, 1975.

16. Tourangeau, R., & Sternberg, R. J. Understanding and appreciating metaphors. NR 150-412 ONR Technical Report No. 11. New Haven: Department of Psychology, Yale University, June, 1978.

17. Tourangeau, R., & Sternberg, R. J. What makes a good metaphor?

Manuscript in preparation, 1978.

18. Frederiksen, J. R. A chronometric study of component skills in reading. Unpublished manuscript, 1978.

19. Shepard, R. N., & Arabie, P. Additive cluster analysis of similarity data. In Proceedings of U.S.-Japan Seminar on Theory, Methods and Applications of Multidimensional Scaling and Related Techniques, San Diego, August, 1975.

20. Arabie, P., & Carroll, J. D. MAPCLUS: A mathematical programming approach to fitting the ADCLUS model. Manuscript submitted for publication, 1978.

References

Arnold, J. B. A multidimensional scaling study of semantic distance. Journal of Experimental Psychology Monograph, 1971, 90, 349-372.

Atkinson, R. C. Optimizing the learning of a second-language vocabulary. Journal of Experimental Psychology, 1972, 96, 124-129.

Campbell, D. T. Recommendations for APA test standards regarding construct, trait, or discriminant validity. American Psychologist, 1960, 15, 546-553.

Carroll, J. D. Spatial, non-spatial, and hybrid models for scaling. Psychometrika, 1976, 41, 439-463.

Cattell, R. B. Abilities: Their structure, growth, and action. Boston: Houghton-Mifflin, 1971.

Clark, H. H. The influence of language in solving three-term series problems. Journal of Experimental Psychology, 1969, 82, 205-215. (a)

Clark, H. H. Linguistic processes in deductive reasoning. Psychological Review, 1969, 76, 387-404. (b)

Darlington, R. B. Multiple regression in psychological research and practice. Psychological Bulletin, 1968, 69, 161-182.

Egan, D. E. Testing based on understanding: Implications from studies of spatial ability. Intelligence, in press.

Guilford, J. B. The nature of human intelligence. New York: McGraw-Hill, 1967.

Guttman, L. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. Psychometrika, 1968, 33, 469-506.

Henley, N. M. A psychological study of the semantics of animal terms. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 176-184.

Hoerl, A. E., & Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 1970, 12, 55-67. (a)

Hoerl, A. E., & Kennard, R. W. Ridge regression: Applications to nonorthogonal problems. Technometrics, 1970, 12, 69-82. (b)

Hoerl, A. E., & Kennard, R. W. Ridge regression iterative estimation of the biasing parameter. Communications in Statistics, 1976, 5, 77-88.

Horn, J. L. On subjectivity in factor analysis. Educational and Psychological Measurement, 1967, 27, 811-820.

Horn, J. L. Organization of abilities and the development of intelligence. Psychological Review, 1968, 75, 242-259.

Horn, J. L., & Cattell, R. B. Refinement and test of the theory of fluid and crystallized intelligence. Journal of Educational Psychology, 1966, 57, 253-270.

Horn, J. L., & Knapp, J. R. On the subjective character of the empirical base of Guilford's Structure-of-Intellect model. Psychological Bulletin, 1973, 80, 33-43.

Humphreys, L. G., & Fleishman, A. Pseudo-orthogonal and other analysis of variance designs involving individual-differences variables. Journal of Educational Psychology, 1974, 66, 464-472.

Humphreys, L. G., Ilgen, D., McGrath, D., & Montanelli, R. Capitalization on chance in rotation of factors. Educational and Psychological Measurement, 1969, 29, 259-271.

Hunt, E., Lunneborg, C., & Lewis, J. What does it mean to be high verbal? Cognitive Psychology, 1975, 7, 194-227.

Jardine, N., & Sibson, R. Mathematical taxonomy. New York: Wiley, 1971.

Jöreskog, K. G. A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 1969, 34, 183-202.

Jöreskog, K. G. A general method for analysis of covariance structures. Biometrika, 1970, 57, 239-251.

Jöreskog, K. G., & Lawley, D. N. New methods in maximum likelihood factor analysis. British Journal of Mathematical and Statistical Psychology,

1968, 21, 85-96.

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 1964, 29, 1-27. (a)

Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. Psychometrika, 1964, 29, 28-42. (b)

Kruskal, J. B., & Shepard, R. N. A nonmetric variety of linear factor analysis. Psychometrika, 1974, 39, 123-157.

Luce, R. D. Individual choice behavior. New York: Wiley, 1959.

Mosteller, F., & Tukey, J. W. Data analysis: Including statistics. In G. Lindzey & E. Aronson (Eds.), Handbook of social psychology (Second Edition, Vol. 2). Reading, Mass.: Addison-Wesley, 1968.

Mosteller, F., & Tukey, J. W. Data analysis and regression. Reading, Mass.: Addison-Wesley, 1977.

Newell, A., & Simon, H. Human problem solving. Englewood Cliffs, N.J.: Prenctice-Hall, 1972.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. The measurement of meaning. Urbana, Illinois: University of Illinois Press, 1957.

Price, B. Ridge regression: Application to nonexperimental data. Psychological Bulletin, 1977, 84, 759-781.

Quinton, G., & Fellows, B. J. 'Perceptual' strategies in the solving of three-term series problems. British Journal of Psychology, 1975, 66, 69-78.

Rumelhart, D. E., & Abrahamson, A. A. A model for analogical reasoning. Cognitive Psychology, 1973, 5, 1-28.

Schönemann, P. H. The generalized solution of the orthogonal procrustes problem. Psychometrika, 1966, 31, 1-16.

Skinner, H. A. Exploring relationships among multiple data sets. Multivariate Behavioral Research, 1977, 12, 199-220.

Skinner, H. A. The art of exploring predictor-criterion relationships. Psychological Bulletin, 1978, 85, 327-337.

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. Psychometrika, 1962, 27, 125-140. (a)

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. Psychometrika, 1962, 27, 219-246. (b)

Shepard, R. N. Attention and the metric structure of the stimulus space. Journal of Mathematical Psychology, 1964, 1, 54-87.

Shepard, R. N. Representation of structure in similarity data: Problems and prospects. Psychometrika, 1974, 39, 373-421.

Sternberg, R. J. Component processes in analogical reasoning. Psychological Review, 1977, 84, 353-378. (a)

Sternberg, R. J. Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. Hillsdale, N.J.: Erlbaum, 1977. (b)

Sternberg, R. J., & Rifkin, B. The development of analogical reasoning processes. Journal of Experimental Child Psychology, in press.

Sternberg, R. J., Tourangeau, R., & Nigro, G. Metaphor, induction, and social policy: The convergence of macroscopic and microscopic views. In A. Ortony (Ed.), Metaphor and thought. New York: Cambridge University Press, in press.

Tucker, L. R. An inter-battery method of factor analysis. Psychometrika, 1958, 23, 111-136.

Tucker, L. R. An inter-battery factor analytic approach to linear predictive systems. Educational and Psychological Measurement, 1973, 33, 77-97.

Wainer, H. Estimating coefficients in linear models: It don't make no nevermind. Psychological Bulletin, 1976, 83, 213-217. (a)

Wainer, H. Robust statistics: A survey and some prescriptions. Journal of Educational Statistics, 1976, 1, 285-312. (b)

Wainer, H. On the sensitivity of regression and regressors. Psychological Bulletin, 1978, 85, 267-273.

Wainer, H., & Thissen, D. When jackknifing fails (or does it?). Psychometrika, 1975, 40, 113-114.

Young, F. W. Nonmetric multidimensional scaling: Recovery of metric information. Psychometrika, 1970, 35, 455-473.

Young, F. W., & Torgerson, W. S. TORSCA, A FORTRAN IV program for Shepard-Kruskal multidimensional scaling analysis. Behavioral Science, 1967, 12, 498.

Footnotes

Preparation of this article was supported by Contract No. N0001478C0025 from the Office of Naval Research to Robert J. Sternberg. Requests for reprints should be sent to Robert J. Sternberg, Department of Psychology, Yale University, Box 11A Yale Station, New Haven, Connecticut 06520.

[1]Applications of the psychometric concepts of validity and reliability to spatial-ability research in particular, and information-processing research in general, are lucidly discussed by Egan (in press).

[2]This linguistic model, based upon one proposed by Clark (1969b), differs from the one described earlier, proposed by Quinton and Fellows (1975) and tested by Sternberg and Weil (Note 11).

[3]I assume here factor analysis of individuals' scores on items or tests (R-analysis).

[4]My view here represents a modification of an earlier view (Sternberg, 1977b), wherein I stated that factor analysis should not be performed at the level of the component.

Technical Reports Presently in this Series

NR 150-412, ONR Contract N0001478C0025

#1. Sternberg, R. J. Intelligence research at the interface between differential and cognitive psychology: Prospects and proposals. January, 1978.

#2. Sternberg, R. J. Isolating the components of intelligence. January, 1978.

#3. Sternberg, R. J., Guyote, M. J., & Turner, M. E. Deductive reasoning. January, 1978.

#4. Sternberg, R. J. Toward a unified componential theory of human reasoning. April, 1978.

#5. Guyote, M. J., & Sternberg, R. J. A transitive-chain theory of syllogistic reasoning. April, 1978.

#6. Sternberg, R. J., & Turner, M. E. Components of syllogistic reasoning. April, 1978.

#7. Sternberg, R. J., Tourangeau, R., & Nigro, G. Metaphor, induction, and social policy: The convergence of macroscopic and microscopic views. April, 1978.

#8. Sternberg, R. J. A proposed resolution of curious conflicts in the literature on linear syllogistic reasoning. June, 1978.

#9. Sternberg, R. J. The nature of mental abilities. June, 1978.

#10. Sternberg, R. J. Psychometrics, mathematical psychology, and cognition: Confessions of a closet psychometrician. June, 1978.

#11. Tourangeau, R., & Sternberg, R. J. Understanding and appreciating metaphors. June, 1978.